1. TITLE OF THE INVENTION

BUFFER MEMORY CONTROL METHOD

2. SCOPE OF CLAIM

A buffer memory control method used in a data processing system that includes a buffer memory and a priority order management unit that manages a priority order of a block unit, the block unit being of a predetermined number that is transferred to the buffer memory, said method comprising:

providing information for managing a validity of block unit information per unit of the block unit information transferred to said buffer memory;

placing, using said priority order management unit, invalid block unit information in order starting with a low priority, based on said information, and providing information specifying that the block unit information is not invalid but the block should be replaced, for the block unit transferred to the buffer memory; and

controlling said priority order management unit to place the block unit for which the information has been provided lower in the priority order.

3. DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention is relates to a buffer memory control method, and in particular, relates to a buffer memory control method that a replacement process, in which block units are transferred to a buffer memory, stores information for indicating that the prescribed information is no longer valid or information indicating that compared to other information, the prescribed information should be replaced although it is valid. As well as controlling block units, to which the former information or the latter information is given to be replaced.

Since in data processing devices that have a buffer memory, the number of block units that can be stored in the buffer memory is limited, the next block unit that may be replaced is controlled to be determined every time there is an access by setting a priority order management unit.

In this case, based on the logic of the Least Recently Used (LRU), the most recently used block unit is given the highest priority and its opposite block unit is controlled as a block unit to be replaced. Also, in other methods which are not based on the LRU logic, the logic for determining a block unit to be replaced conventionally follows a passive logic where a block unit to be replaced is determined through

comparison with other block units. Contrary to a passive logic, an active logic simply indicates an obviously invalid block unit.

A goal of the present invention is to indicate a block unit that is not invalid as information but should be actively replaced, and to actively replace such block unit due to the priority order management unit. To this end, the buffer control method of the present invention is a buffer memory control method used in a data processing system that includes a buffer memory and a priority order management unit that manages a priority order of a block unit, the block unit being of a predetermined number that is transferred to the buffer memory, and includes: providing information for managing a validity of block unit information per unit of the block unit information transferred to the buffer memory; placing, using the priority order management unit, invalid block unit information in order starting with a low priority, based on the information, and providing information specifying that the block unit information is not invalid but the block should be replaced, for the block unit transferred to the buffer memory; and controlling the priority order management unit to place the block unit for which the information has been provided lower in the priority order. Description shall be given with reference to the figures.

Figure 1 shows an example of a data processing system in which the buffer memory control method of the present invention is applied. Figure 2 shows an example structure of address information for assigning specific memory information. Figure 3 shows an example structure of the block unit replacement determination in the present invention.

In Figure 1, each number corresponds as follows: 1 is a main memory; 2 is a central processing unit having a buffer memory; 3 is a buffer memory; 4 is a directory for managing the block unit address information stored in 3; 5 is a memory access device; 6 is an instruction control unit; 7 is a processor; 8-0 and 8-1 are channel control devices; 9 is a file control device; 10 is a mass external storage; 11 is an input/output control device; 12 is an input/output device.

The main memory 1 is segmented into 8 block unit 13 and has a total number of block units of m (tag) x n (column). 4 block units per column can be transferred to and stored in the buffer memory 3. When it is necessary to access information within a block unit that is not present in the buffer memory 3, the block unit with the lowest priority of the 4 block units is controlled to be replaced. That is, corresponding to access by the access origin, such as the central processing unit 2 and the channel control 8-0 or 8-1, the directory 4 is searched for necessary information; if the necessary information is present in the buffer memory 3, such information is read out from the buffer memory

3. If such information is not present, the main memory 1 is accessed for the necessary information, which is transferred to the access origin. At the same time, the block unit including such information is stored in the buffer memory 3. The lowest priority block unit of the four block units in the buffer memory 3 is extracted.

In this case, LRU is traditionally employed as the logic for determining a block to be replaced,. Also, when a block-unit information in the buffer memory 3 becomes invalid due to being rewritten by another device in the main memory 1, this invalid block unit is replaced.

In addition to the conventional replacement processing logic, the present invention makes it possible to specify a block unit to be actively replaced. In other words,

(1) As shown in Figure 1, in the case where the channel control 8-0 transfers data from main memory to the mass external storage 10, the block unit information including such data is stored in the buffer memory 3. However, after the last data within the block unit is accessed, there is no longer any value in the block unit being present in the buffer memory 3.

(2) When comparing to information accessed by the central processing unit 2 with information accessed by the channel control 8-0 or 8-1, in the former case, adjacent addresses are more likely to be accessed by turns according to the implementation of processes, while in the latter case, because adjacent addresses are less likely to be accessed, after a block unit is accessed, placing the unit in the buffer memory 3 is of less value.

(3) Considering access by the access origin, for example, it is of comparatively less value to place data in the buffer memory 3 when the access origin accesses the user domain.

(4) Considering a monitor mode and a user mode of a program in the data processing unit, the latter is of less value to be placed in the buffer memory 3.

(5) Depending on the types of processing of the data processing unit, it can be expected that there is no or extremely little possibility that after certain information is accessed, such information is accessed again for a certain amount of time. In this case, certain block units have no longer any value to be present in the buffer memory 3 after they have been accessed once.

In such cases, block units being of less value are controlled to be actively replaced automatically or through an instruction.

As shown in Figure 2, the address information in the present invention is specified in "tag address information", "column address information", and

"address-within-block information".

In figure 3, each number shows: 4 is a directory; 13 is an address register; 14-0, 14-1, 14-2 and 14-3 are comparators; 15 is a priority order determination circuit; 16-0, 16-1, 16-2 and 16-3 are the lowest order set flip flop.

As explained in connection with Figure 1, when the buffer memory 3 is accessed, address information is set in the address register 13 of access origin, and the directory 4 is read out based on the column address information. For example, if such column address information is corresponding to the # 0 column, the tag address information is read out all at once from four set positions of the # 0 column of the directory 4, that is the # 0 set position to the # 3 set position. This tag address information that has been read out is the same tag address information per block stored in four set positions of the # 0 column in the buffer memory 3 (Figure 1) and is compared with the tag address information in the register 13 using the comparators 14-0 to 14-3. Then, if mismatch signals are emitted from all comparators, the corresponding set position in the buffer memory 3 or a block unit stored in the # 1 set position is expelled and a new block unit is transferred from the main memory 1 to that position. Then, the # 1 set position, in which a new block unit is stored, is given the highest priority, and after the lowest priority is determined, a relevant flip flop such as 16-2 is set.

In this case, the information V, which indicates whether or not a block unit stored in the buffer memory 3 is valid, is stored together in the directory 4. To determine the lowest priority, if there is a block unit where the information V is logic "0" (that is invalid), the flip flop of the existing set position of such block unit is set.

In the present invention, in addition to the information V and the process of determining the priority order, information indicating that " not invalid as information but should be replaced" illustrated as the information R in the directory 4 is controlled to be stored (not limited to 1 bit). Also, a block unit in which such information R is logic "1" (that is it should be promptly replaced) is actively replaced. In other words, firstly, the block where the information V is at logic 0 placed at the lowest priority. Secondly, when all the information V is at logic 1, the block where the information R is at logic 1 is placed at the lowest priority. Thirdly, when the information V is at logic 1 and the information R is at logic 0 for all blocks, for example, the lowest priority block is controlled to be determined based on the LRU logic.

A block unit to be specified as logic 1 on the information R is determined according to the above mentioned (1) to (5) and so on. Then, the information R is automatically changed to logic 1 or is changed to logic 1 through an instruction.

As explained above, according to the present invention, in the process of determining a block to be replaced through the priority order determination circuit 15, a block to be actively replaced is processed to be easily replaced according to the information R. Compared to the conventional method which equally considers all block units that are not invalid and employs logic such as the LRU logic, the present invention allows the implementation of a more effective buffer memory control.

4. BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows an example of the data processing system to which the buffer memory control method of the present invention is applied. Figure 2 shows an example structure of the address information of assigning specific memory information. Figure 3 shows an example structure of determining a block unit to be replaced according to the present invention.

The figures show: 1 is a main memory; 2 is a central processing unit; 3 is a buffer memory; 4 is a directory; 8-0 and 8-1 are channel control devices; 9 is a file control device; 10 is a mass external storage; 15 is a the priority order determination circuit; V is information to manage information validity; R is information to indicate that information is not invalid but should be replaced

Patent Applicant: Fujitsu Limited
Patent Agent: Hiroshi Morita

Figure 1 — 第1図

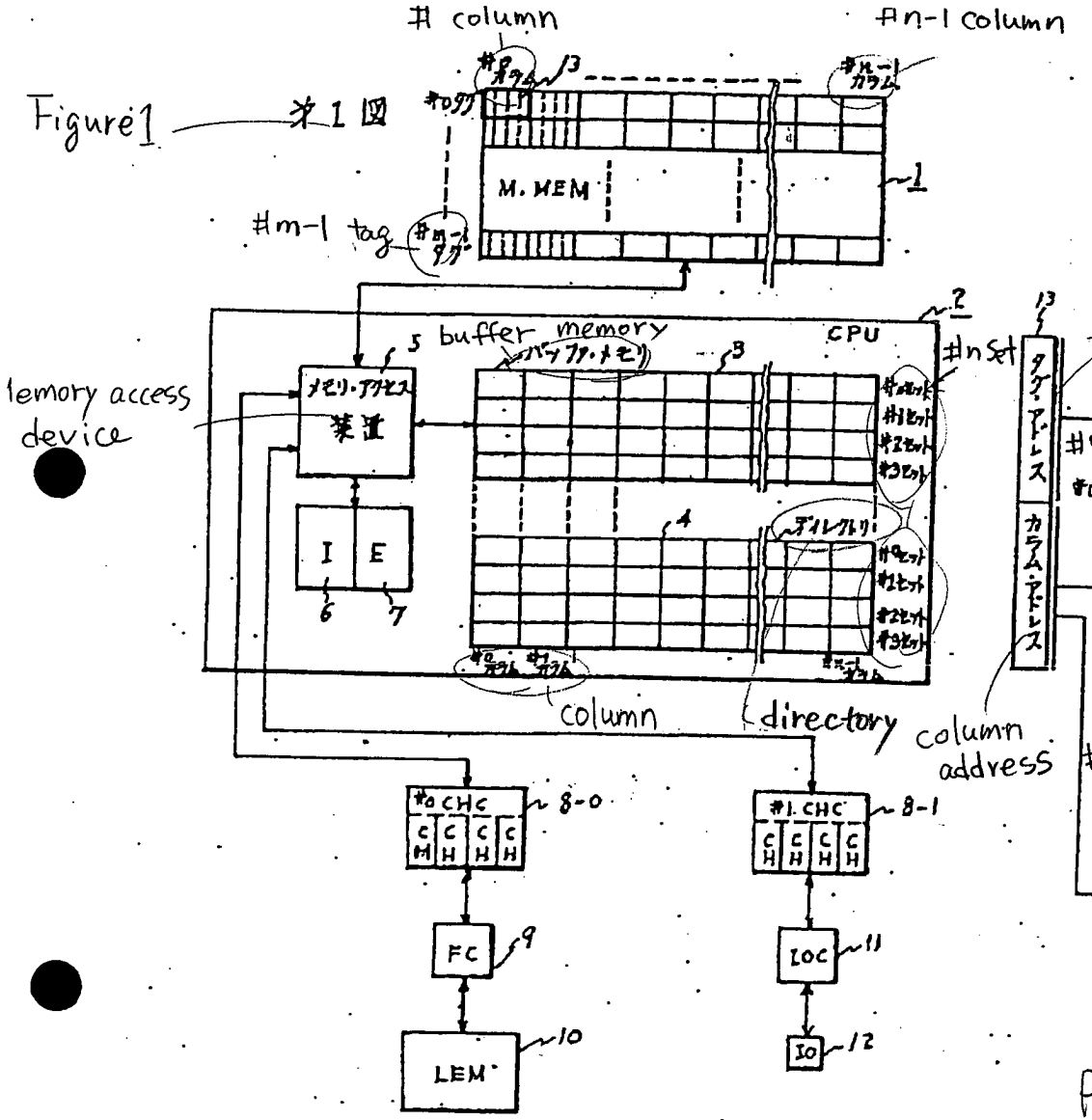


Figure 2 — 第2図

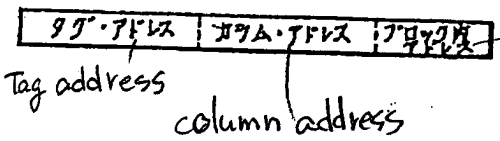
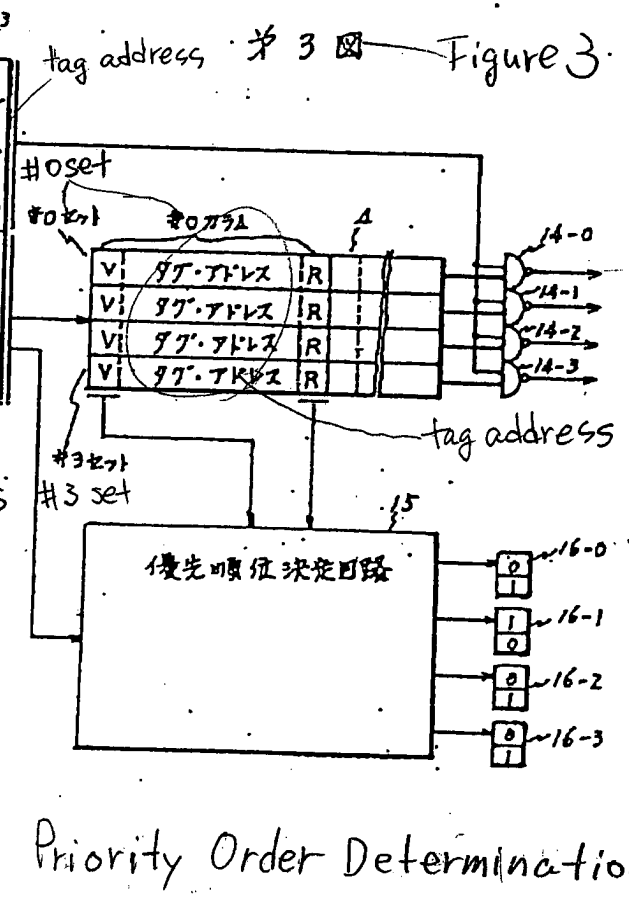


Figure 3 — 第3図



Priority Order Determination Circuit

⑲ 日本国特許庁

公開特許公報

⑪特開昭 51-19453

㊸公開日 昭51.(1976) 2.16

㉑特願昭 49-91075

㉒出願日 昭49.(1974) 8.8

審査請求 未請求 (全5頁)

庁内整理番号

6453 56

6453 56

㊷日本分類	㊶Int.Cl²
97(7)C0 97(7)C02	G06F 13/00 G11C 9/00

特 許 願

昭和49年8月8日

特許庁長官 斉藤英雄殿

1. 発明の名称 バッファ・メモリ制御方式

2. 発明者

住所 神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

氏名 田村 愛 (外2名)

3. 特許出願人

住所 神奈川県川崎市中原区上小田中1015番地

氏名 (522) 富士通株式会社

代表者 高嶺芳光

4. 代理人

住所 東京都荒川区西日暮里4丁目17番1号
佐原マンション3FC

氏名 (7484) 弁理士 森田 寛

5. 添付書類の目録

(1)	明細書	1通
(2)	図面	1通
(3)	委任状	1通
(4)	願書副本	1通

方式審査

明細書

1. 発明の名称 バッファ・メモリ制御方式

2. 特許請求の範囲

バッファ・メモリ，および該バッファ・メモリ上に転送されている所定個数のブロック単位について該ブロック単位の優先順位を管理する優先順位管理手段をそなえたデータ処理システムにおいて，上記バッファ・メモリに転送されている各ブロック単位情報について該ブロック単位情報の有効性を管理する情報をもうけ，該情報にもとずいて無効ブロック単位情報を上記優先順位管理手段によりもっとも低い優先順位におくと共に，上記バッファ・メモリに転送されているブロック単位について該ブロック単位情報が無効ではないが置換されるべきである旨を指示する情報をもうけ，該情報が附与されているブロック単位を，上記優先順位管理手段により，より低い優先順位におくよう制御せしめたことを特徴とするバッファ・メモリ制御方式。

3. 発明の詳細な説明

本発明は，バッファ・メモリ制御方式，特にバッファ・メモリに転送されているブロック単位の置換処理において，所定ブロック単位情報がもはや有効でないことを指示する情報，および有効ではあるが他にくらべて置換されるべき旨を指示する情報を保持しておくようにし，上記前者情報または後者情報が附与されているブロック単位が積極的に置換され易いよう制御せしめたバッファ・メモリ制御方式に関するものである。

バッファ・メモリを有するデータ処理装置においては，バッファ・メモリ上に格納できるブロック単位数に制限があることから，優先順位管理手段をもうけてアクセスがある度に次に置換される可能性のあるブロック単位を決定しておくように制御される。

この場合従来からLRU(リースト・リーセントリ・ユースド)の論理にしたがって，最も最近

れるが，当該ブロック単位内の最後のデータがアクセスされた後を考えると，当該ブロック単位はもはやバッファ・メモリ3上に存在する価値がなくなる。

(2) 中央処理装置2がアクセスした情報とチャネル制御装置8－0または8－1がアクセスした情報とを対比して考えると，前者の場合処理の進行に応じて隣接したアドレスが順にアクセスされる可能性が大であるのに対し，後者の場合その可能性が小であることから，後者がアクセスした後のブロック単位をバッファ・メモリ3上に置く価値が少ない。

(3) アクセス元によるアクセスを考えたとき例えばユーザ領域に相当するアクセスは他にくらべてバッファ・メモリ3上に置かれる価値が少ない。

(4) データ処理装置におけるプログラムについてモニタ・モードとユーザ・モードとを考えて，後者はバッファ・メモリ3上に置かれる価値が少ない。

(5) データ処理装置の処理の種類によっては，ある情報をアクセスした後当分の間再びその情報を含むブロック単位についてアクセスされる可能性がないかあるいはきわめて少ないことが予測できる場合があり，このような場合特定のブロック単位は一度アクセスされた後にはもはやバッファ・メモリ3上に存在する価値がなくなる。

などの各種ケースにおいて，バッファ・メモリ3上に存在する価値がないか少ないブロック単位に対して，自動的にあるいは命令によって，積極的に置換されるべきものとして指定するようにする。

本発明におけるアドレス情報は第2図に表わす如く，（タグ・アドレス情報）・（カラム・アドレス情報）・（ブロック内アドレス情報）で指定される。

第3図において，4はディレクトリ，13はアドレス・レジスタ，14－0，14－1，14－2，14－3は比較回路，15は優先順位決定回路，16－0，16－1，16－2，16－3は最下位順位セット・フリップ・フロップを表わしている。

第1図に関連して説明した如く，バッファ・メモリ3に対してアクセスする場合，アクセス元はアドレス・レジスタ13にアドレス情報がセットされ，ディレクトリ4に対してカラム・アドレス情報にもとずいて読出し処理が行なわれる。そして該カラム・アドレス情報が例えば#0カラムに相当するものであった場合，ディレクトリ4の#0カラムの4つのセット位置#0セット位置ないし#3セット位置からタグ・アドレス情報が一斉に読出される。この読出されたタグ・アドレス情報は，バッファ・メモリ3（第1図）の#0カラムの4つのセット位置に格納されているブロック単位のタグ・アドレス情報であって，比較回路14－0ないし14－3によってレジスタ13上のタグ・アドレス情報と比較される。そしてすべての比較回路から不一致信号が発生られると，セット状態にあるフリップ・フロップ例えば16－1によってバッファ・メモリ3上の対応するセット位置即ち#1セット位置に格納されているブロック単位が追出され，その位置に新しいブロック単位が主記憶装置1から転送されてくる。そして新しいブロック単位が格納された#1セット位置を最高の優先順位に置くようにし，新しく最下位の順位が決定されて該当するフリップ・フロップ例えば16－2がセットされる。

そしてこの場合，ディレクトリ4内にはバッファ・メモリ3上に格納されているブロック単位が有効な情報であるか否かを指示する情報Vが一緒に格納されており，上記最下位優先順位の決定には情報Vが論理「0」（即ち無効）となっているブロック単位が存在すれば，当該ブロック単位の存在するセット位置のフリップ・フロップをセットしておくようにされる。

本発明においては，上記情報Vおよび例えばLRU論理による優先順位決定処理に加えて，ディレクトリ4上に情報Rとして図示（1ビットに限らない）した「情報として無効ではないが置換されるべき旨」を指示する情報を格納せしめておく

6. 前記以外の発明者

発明者

住所　神奈川県川崎市中原区上小田中1015番地
富士通株式会社内

氏名　矢澤正芝

住所　同上

氏名　下村勝弥